Sampling and Stability

C. Rieger, R. Schaback, B. Zwicknagl

Abstract. In Numerical Analysis one often has to conclude that an error function is small everywhere if it is small on a large discrete point set and if there is a bound on a derivative. *Sampling inequalities* put this onto a solid mathematical basis.

A *stability inequality* is similar, but holds only on a finite–dimensional space of trial functions. It allows to bound a trial function by a norm on a sufficiently fine data sample, without any bound on a high derivative.

This survey first describes these two types of inequalities in general and shows how to derive a stability inequality from a sampling inequality plus an inverse inequality on a finite–dimensional trial space. Then the state–of–the–art in sampling inequalities is reviewed, and new extensions involving functions of infinite smoothness and sampling operators using weak data are presented.

Finally, typical applications of sampling and stability inequalities for recovery of functions from scattered weak or strong data are surveyed. These include Support Vector Machines and unsymmetric methods for solving partial differential equations.

§1. Introduction

In many practical applications it is necessary to approximate or reconstruct a function as a formula from given strong or weak scattered data. Important examples are domain modeling, surface reconstruction, machine learning or the numerical solution of partial differential equations.

If the strong or weak data used for reconstruction are seen as a *sampling* $S(f) \in \mathbb{R}^N$ of an unknown function f, and if a *trial* function u satisfying $S(u) \approx S(f)$ is calculated in order to recover f, then $S(f - u)$ is small and one has to conclude that $f - u$ is small. Of course, such a conclusion requires additional assumptions, e.g., a bound on derivatives of $f - u$.

The *sampling inequalities* surveyed here quantify the observation that a differentiable function cannot attain large values anywhere if its derivatives

are bounded, and if it produces small data on a sufficiently dense discrete set. Along the lines of the above argument, such inequalities are extremely useful to derive *a priori* error estimates for very general approximation processes [30].

In the univariate setting, sampling inequalities are quite easy to obtain.

We assume a sufficiently smooth function f on an interval $[a, b]$ and a discrete ordered set of centers $X = \{x_1, \dots, x_N\} \subset [a, b]$ with

$$a = x_1 < x_2 < \cdots < x_{N-1} < x_N = b$$

and the *fill distance*

$$h := h_{X,[a,b]} := \frac{1}{2} \max_{2 \leq j \leq N} |x_j - x_{j-1}|,$$

i.e., the largest possible distance any point $x \in [a, b]$ has from the set X. With this notation we consider an arbitrary point $x \in [a, b]$ and its closest point $x_j \in X$ to get

$$f(x) = f(x_j) + \int_{x_j}^{x} f'(t)\, dt,$$

$$|f(x)| \leq |f(x_j)| + \sqrt{|x - x_j|} \sqrt{\int_{x_j}^{x} |f'(t)|^2\, dt},$$

which yields a first instance of a *sampling inequality*

$$\|f\|_{L_\infty([a,b])} \leq \sqrt{h} |f|_{W_2^1[a,b]} + \|S_X(f)\|_{\ell_\infty(\mathbb{R}^N)}$$

for the *sampling operator*

$$S_X \ : \ W_2^1[a, b] \to \mathbb{R}^N \text{ with } S_X(f) := (f(x_1), \dots, f(x_N))^T. \qquad (1)$$

This easy example already reveals the basic phenomenon, i.e., it bounds a function in a weak continuous norm in terms of the sampled data on a discrete set X and a strong continuous norm weighted by a power of the fill distance of X. We shall explain this in general in the next section, while we postpone specific applications to sections 3 and 4.

§2. Sampling and Stability

Here, we shall exhibit general features of *sampling* and *stability* inequalities and their connections. We admit general spaces of multivariate functions and general sampling operators. Specific cases will follow in section 3.

2.1. Sampling Inequalities

We consider a linear space $\mathcal{F}$ of real–valued functions on some domain $\Omega \subset \mathbb{R}^d$, and assume $\mathcal{F}$ to carry norms $\|\cdot\|_{\mathfrak{S}}$ and $\|\cdot\|_W$, where $\|\cdot\|_{\mathfrak{S}}$ is stronger than $\|\cdot\|_W$, i.e.,

$$\|f\|_W \le C \|f\|_{\mathfrak{S}} \quad \text{for all } f \in \mathcal{F}. \tag{2}$$

Here and in the following, C denotes a generic positive constant which is independent of the terms in the following *for all* statement, but sometimes we add dependencies of constants on certain problem parameters by adding argument lists. For example, we could have written $C(\mathcal{F}, \mathfrak{S}, W)$ in the inequality (2).

Furthermore, we consider classes $\mathcal{L}$ of finite sets $\Lambda = \{\lambda_1, \ldots, \lambda_N\}$ of linearly independent functionals from the dual space $\mathcal{F}^*$ with respect to $\|\cdot\|_{\mathfrak{S}}$. These functionals are used to *sample* a function from $\mathcal{F}$ via the continuous and linear *sampling operator*

$$\begin{aligned} S_\Lambda \quad &: \quad \mathcal{F} \to \mathbb{R}^N, \\ f \quad &\mapsto (\lambda_1(f), \ldots, \lambda_N(f))^T, \end{aligned}$$

generalizing (1). The term h_Λ denotes some discretization parameter which should be small, i.e., $h_\Lambda \overset{N \to \infty}{\longrightarrow} 0$. Then an abstract form of a sampling inequality is

$$\|f\|_W \le C\left(h_\Lambda^\sigma \|f\|_{\mathfrak{S}} + C(h_\Lambda) \|S_\Lambda(f)\|_{\mathbb{R}^N}\right) \quad \text{for all } f \in \mathcal{F}, \Lambda \in \mathcal{L}. \tag{3}$$

Sometimes such an inequality holds even if $\|\cdot\|_{\mathfrak{S}}$ is replaced by a semi-norm $|\cdot|_{\mathfrak{S}}$ with finite dimensional kernel $\mathcal{P}_{\mathfrak{S}}$, and then we rewrite this as

$$\|f\|_W \le C\left(h_\Lambda^\sigma |f|_{\mathfrak{S}} + C(h_\Lambda) \|S_\Lambda(f)\|_{\mathbb{R}^N}\right) \quad \text{for all } f \in \mathcal{F}, \Lambda \in \mathcal{L}. \tag{4}$$

The exponent $\sigma > 0$ is called *sampling order*. Hence there is a small factor in front of the term with the strong continuous norm and a possibly large term in front of the term with the discrete norm. Furthermore, in most cases the class $\mathcal{L}$ of admissible samplings must be "sufficiently fine" in the sense that $h_\Lambda \le h_0$ holds for some positive constant h_0.

If the sampling operator contains only point evaluations based on a finite point set $X = \{x_1, \ldots, x_N\} \subset \Omega$, we write it as

$$S_X := (\delta_{x_1}, \ldots, \delta_{x_N})^T \tag{5}$$

like in (1), and the discretization parameter is then given by the *fill distance*

$$h_{X,\Omega} := \sup_{y \in \Omega} \max_{x \in X} \|x - y\|_2 \tag{6}$$

of the discrete set X with respect to the domain Ω. Geometrically, the fill distance $h_{X,\Omega}$ can be interpreted as the radius of the largest open ball with center in the closure $\overline{\Omega}$ that does not contain any of the points from X. It is a useful quantity for the deterministic error analysis in Sobolev spaces, i.e., if there are no further structural assumptions on the approximated functions. On the other hand, the *separation distance q_X* defined by

$$q_X := \frac{1}{2} \min_{j \neq k} \|x_j - x_k\|_2 \tag{7}$$

is the largest radius of balls with centers in X that do not contain any other of the points from X, and it is crucial for stability questions.

If the sampling operators of the class $\mathcal{L}$ consist of evaluations of f and its derivatives on certain finite point sets $X \subset \Omega$, we speak of *strong sampling inequalities*. If some other functionals are involved, which may be well defined even if point evaluation is not continuous, we speak of *weak sampling inequalities*. We shall treat strong and weak sampling separately from section 3 on.

2.2. Special Stability Inequalities

Continuing the notation of (4), we denote the kernel of the semi-norm $|\cdot|_{\mathfrak{S}}$ by $\mathcal{P}_{\mathfrak{S}} \subset \mathcal{F}$. If we insert an element $p \in \mathcal{P}_{\mathfrak{S}}$ from this kernel into the sampling inequality (4), we obtain

$$\|p\|_W \leq C \left(h_\Lambda^\sigma |p|_{\mathfrak{S}} + C (h_\Lambda) \|S_\Lambda (p)\|_{\mathbb{R}^N} \right) = C (h_\Lambda) \|S_\Lambda (p)\|_{\mathbb{R}^N} \tag{8}$$

for all $p \in \mathcal{P}_{\mathfrak{S}}$ and $\Lambda \in \mathcal{L}$. This means that we can bound a continuous norm by a discrete norm on the data. Bounds of this kind will be called *stability inequalities*. They follow from sampling inequalities and hold on the kernel $\mathcal{P}_{\mathfrak{S}}$ of the strong semi–norm involved, if it is finite-dimensional. If $\mathcal{P}_{\mathfrak{S}}$ is a space of polynomials, and if $S_\Lambda = S_X$ is a strong pointwise sampling operator (5) on a finite set X, these estimates imply Markov-Bernstein inequalities [13]. Let us explain this in some more detail. Assume that $|\cdot|_W = |\cdot|_{W_\infty^1(\Omega)}$, and that $|\cdot|_{\mathfrak{S}} = |\cdot|_{W_\infty^k(\Omega)}$ for $k > 1$ are classical Sobolev semi-norms. This yields for all $1 \leq \ell \leq d$

$$\begin{aligned}
\|\partial_\ell p\|_{L_\infty(\Omega)} &\leq C (h_\Lambda) \|S_X (p)\|_{\ell_\infty(\mathbb{R}^N)} \\
&\leq C (h_\Lambda) \|p\|_{L_\infty(\Omega)} \text{ for all } p \in \pi_{k-1}(\Omega),
\end{aligned} \tag{9}$$

where ∂_ℓ denotes the partial derivative in direction of the ℓ-th coordinate. This is a special case of classical Markov-Bernstein-inequalities [13]. Thus it is not surprising that the proofs for sampling inequalities use those classical estimates. The stability inequality (8) implies that the data $S_\Lambda (p)$ contains already enough information about $p \in \mathcal{P}_{\mathfrak{S}}$. This is connected to the general concept of *norming sets* [20]. We shall briefly explain this concept, since it is a direct way for proving stability inequalities under certain

circumstances. A finite set Λ of linear functionals is called a *norming set* for $\mathcal{P}_{\mathfrak{S}}$ if the sampling operator

$$
\begin{aligned}
S_\Lambda|_{\mathcal{P}_{\mathfrak{S}}} : \quad \mathcal{P}_{\mathfrak{S}} &\to S_\Lambda|_{\mathcal{P}_{\mathfrak{S}}} (\mathcal{P}_{\mathfrak{S}}) \subset \mathbb{R}^N \\
v &\mapsto S_\Lambda|_{\mathcal{P}_{\mathfrak{S}}} (v) = (\lambda(v))_{\lambda \in \Lambda}
\end{aligned}
$$

is injective. Then we can introduce another norm on $\mathcal{P}_{\mathfrak{S}}$ by $\|S_\Lambda|_{\mathcal{P}_{\mathfrak{S}}} (\cdot)\|_{\mathbb{R}^N}$ and we immediately get a stability inequality (8). This explains the terminology *norming set*. We shall explain below why norming sets are crucial in the proofs of sampling inequalities.

2.3. Inverse Inequalities and General Stability Inequalities

Stability inequalities like (8) are limited to the kernel $\mathcal{P}_{\mathfrak{S}}$ of the strong semi-norm in (4). But they should generalize to finite–dimensional trial spaces $\mathcal{R}$ in the sense

$$
\|u\|_W \le C_{stab}(\Lambda, \mathcal{R}, W)\|S_\Lambda u\|_{\mathbb{R}^N} \text{ for all } u \in \mathcal{R} \tag{10}
$$

bounding a weak continuous norm by a discrete norm. This is obvious due to the fact that all norms on finite–dimensional spaces are equivalent, but it is hard to determine the constants $C(\Lambda, \mathcal{R}, W)$. Clearly, (10) implies that the sampling operator $S_\Lambda|_{\mathcal{R}} : \mathcal{R} \to \mathbb{R}^N$ is injective, i.e., that Λ is a norming set for $\mathcal{R}$.

To let a sampling inequality produce such a stability inequality, we could use (4) on $\mathcal{R}$ to get

$$
\|u\|_W \le C \left(h_\Lambda^\sigma |u|_{\mathfrak{S}} + C(h_\Lambda) \|S_\Lambda(u)\|_{\mathbb{R}^N} \right) \quad \text{for all } u \in \mathcal{R}. \tag{11}
$$

The second part of the right–hand side is what we want, but the first part should go away, preferably by moving it to the left–hand side. Thus we want to bound a strong semi-norm by a weaker norm on the trial space like

$$
|u|_{\mathfrak{S}} \le C_{inv}(\mathcal{R}, W, \mathfrak{S}) \|u\|_W \text{ for all } u \in \mathcal{R}, \tag{12}
$$

and this is called an *inverse inequality*. We point out, that such an inequality can hold only if the constant $C_{inv}(\mathcal{R}, W, \mathfrak{S})$ grows to infinity if the trial discretization becomes finer and finer. This inequality, however, is independent of the sampling, and together with our sampling inequality (11) it provides the stability inequality (10) with

$$
C_{stab}(\Lambda, \mathcal{R}, W) = 2CC(h_\Lambda)
$$

provided that we have the *stability condition*

$$
Ch_\Lambda^\sigma C_{inv}(\mathcal{R}, W, \mathfrak{S}) < \frac{1}{2}, \tag{13}
$$

which can always be satisfied if h_Λ is sufficiently small, i.e., the sampling is "fine enough". This means that in general

$$
\left.\begin{array}{c}
\text{Sampling Inequality} \\
+ \\
\text{Inverse Inequality}
\end{array}\right\} \Rightarrow \text{Stability Inequality.}
$$

2.4. Connection to Lebesgue Constants

Sampling and stability inequalities are closely related to Lebesgue constants. To see this, assume a trial space $\mathcal{R}$ that allows unique generalized interpolation for N linearly independent functionals $\lambda_1, \ldots, \lambda_N$ defining a sampling operator S_Λ with values in $\mathbb{R}^N$. Then we can build the generalized cardinal interpolants u_{λ_i} from $\mathcal{R}$ with $\lambda_j(u_{\lambda_i}) = \delta_{i,j}$, where $\delta_{i,j}$ denotes the Kronecker symbol, and the interpolant to a function f is $I_\Lambda(f)(\cdot) = \sum_{j=1}^N \lambda_j(f) u_{\lambda_j}(\cdot)$. This yields a stability estimate of the form

$$
\begin{aligned}
\|I_\Lambda(f)\|_W &= \left\| \sum_{j=1}^N \lambda_j(f) u_{\lambda_j} \right\|_W \leq \sum_{j=1}^N |\lambda_j(f)| \, \|u_{\lambda_j}\|_W \\
&\leq \max_{1 \leq j \leq N} |\lambda_j(f)| \sum_{j=1}^N \|u_{\lambda_j}\|_W = \|S_\Lambda(f)\|_{\ell_\infty(\mathbb{R}^N)} L(\mathcal{R}, \Lambda, W)
\end{aligned}
$$

for all $f \in \mathcal{F}$, where the *Lebesgue constant* is defined by

$$
L(\mathcal{R}, \Lambda, W) := \sum_{j=1}^N \|u_{\lambda_j}\|_W .
$$

Now we explain how sampling inequalities can lead to bounds on Lebesgue constants under suitable conditions. If we measure the discrete term in the $\ell_\infty(\mathbb{R}^N)$ norm, i.e.,

$$
\|S_\Lambda(u)\|_{\ell_\infty(\mathbb{R}^N)} = \max_{1 \leq i \leq N} |\lambda_i(u)|,
$$

we immediately get $\|S_\Lambda(u_{\lambda_i})\|_{\ell_\infty(\mathbb{R}^N)} = 1$ for all $1 \leq i \leq N$. Applying the sampling inequality (4) yields

$$
\begin{aligned}
\|u_{\lambda_i}\|_W &\leq C_1 \left(h_\Lambda^\sigma |u_{\lambda_i}|_{\mathfrak{S}} + C_2(h_\Lambda) \|S_\Lambda(u_{\lambda_i})\|_{\ell_\infty(\mathbb{R}^N)} \right) \\
&\leq C_1 \left(h_\Lambda^\sigma |u_{\lambda_i}|_{\mathfrak{S}} + C_2(h_\Lambda) \right).
\end{aligned}
$$

Since we have just one sampling operator, we cannot make sure that a stability condition like (13) holds, but in certain situations (see [17] for

the case of pointwise interpolation by translates of positive definite kernels) there may be bounds of the form

$$|u_{\lambda_i}|_{\mathfrak{S}} \leq C h_{\Lambda}^{-\sigma} \text{ for all } i, \ 1 \leq i \leq N$$

and one has $C_2 (h_{\lambda}) = 1$. This leads to boundedness of the norms $\|u_{\lambda_i}\|_W$, and consequently, by the Cauchy-Schwarz inequality, the Lebesgue constant is bounded above by $\mathcal{O}\left(\sqrt{N}\right)$.

§3. Variations of Sampling Inequalities

After describing sampling inequalities in general together with their close connection to stability, we now turn to special cases involving spaces of multivariate functions. We review the first sampling inequalities dating back to 2005 to 2007 and give a proof sketch. Then we turn to recent extensions to functions with unlimited smoothness and weak sampling operators.

Throughout this survey we shall deal with a variety of Sobolev spaces defined as in [14] and based on a bounded Lipschitz domain Ω satisfying an interior cone condition.

3.1. Strong Sampling Inequalities: Finite Smoothness

Using the notation of the previous sections in the special case $\|\cdot\|_{\mathfrak{S}} = \|\cdot\|_{W_p^k(\Omega)}$ and $\|\cdot\|_W = \|\cdot\|_{W_q^m(\Omega)}$, the condition (2) reduces to

$$W_q^k (\Omega) \hookrightarrow W_p^m (\Omega),$$

i.e., by Sobolev's embedding theorem $k \geq m$ and $k - \frac{d}{q} \geq m - \frac{d}{p}$. Since 2005, several *strong sampling inequalities* for functions $u \in W_p^k (\Omega)$ from Sobolev spaces $W_p^k (\Omega)$ with $1 < p < \infty$ and $k > d/p$, or with $p = 1$ and $k \geq d$ on domains $\Omega \subset \mathbb{R}^d$ have been obtained. As a first step in this direction, Narcowich, Ward and Wendland considered the case of functions with scattered zeros [24]. They proved the existence of positive constants C and h_0 such that the inequality

$$|u|_{W_q^m(\Omega)} \leq C h^{k-m-d\left(\frac{1}{p}-\frac{1}{q}\right)_+} |u|_{W_p^k(\Omega)}$$

holds for all functions $u \in W_p^k (\Omega)$ with $k - m > d/p$ and $S_X (u) = 0$ on arbitrary discrete sets X whose fill distance h in the sense of (6) satisfies $h \leq h_0$. The constants C, h_0 may depend on q, m, p, k, Ω, and d, but not on X, h or u. In [37] this result was generalized to functions with arbitrary values on scattered locations:

Theorem 1. *We assume $1 \leq q \leq \infty$, $\alpha \in \mathbb{N}_0^d$, $k \in \mathbb{N}$, and $1 \leq p < \infty$ with $k > |\alpha| + d/p$ if $p > 1$, or with $k \geq |\alpha| + d$ if $p = 1$. Then there are constants C, $h_0 > 0$ such that*

$$\|D^\alpha u\|_{L_q(\Omega)} \leq C \left(h^{k-|\alpha|-d\left(\frac{1}{p}-\frac{1}{q}\right)_+} |u|_{W_p^k(\Omega)} + h^{-|\alpha|}\|S_X u\|_{\ell_\infty(\mathbb{R}^N)} \right)$$

holds for all $u \in W_p^k(\Omega)$ and all discrete sets $X \subset \Omega$ with sampling operators S_X from (5) and fill distance $h := h_{X,\Omega} \leq h_0$.

A similar result was established by Madych [22] in 2006, namely

$$\|u\|_{L_p(\Omega)} \leq C \left(h^k |u|_{W_p^k(\Omega)} + h^{d/p}\|S_X u\|_{\ell_p} \right) \tag{14}$$

for all $u \in W_p^k(\Omega)$ and all X with $h_{X,\Omega} < h_0$. Arcangéli et al. [11] generalized these sampling inequalities by greatly extending the range of parameters:

Theorem 2. *[11, Thm. 4.1] Let p, q, $\kappa \in [1, \infty]$, and let $r \in \mathbb{R}$ such that $r \geq n$ if $p = 1$, $r > n/p$ if $1 < p < \infty$, or $r \in \mathcal{N}^*$ if $p = \infty$. Likewise, let $i_0 = r - n \left(\frac{1}{p} - \frac{1}{q} \right)_+$ and $\gamma = \max\{p, q, \kappa\}$. Then, there exist two positive constants h_0 and C satisfying the following property: for any finite set $X \subset \bar{\Omega}$ (or $X \subset \Omega$ if $p = 1$ and $n = r$) with fill distance $h := h_{X,\Omega} \leq h_0$, for any $u \in W_p^r(\Omega)$ and for any $l = 0, \ldots, \lceil l_o \rceil - 1$, we have*

$$|u|_{W_q^l(\Omega)} \leq C \left(h^{r-l-n\left(\frac{1}{p}-\frac{1}{q}\right)_+} |u|_{W_p^r(\Omega)} + h^{\frac{n}{\gamma}-l} \|u|_X\|_{\ell_\infty(X)} \right).$$

If $r \in \mathbb{N}^*$ this bound also holds with $l = l_0$ if either $p < q < \infty$ and $l_0 \in \mathbb{N}$, or $(p, q) = (1, \infty)$, or $p \geq q$.

There are several variations, extensions and applications of such sampling inequalities, including derivative data, inequalities on unbounded domains and applications to spline interpolation and smoothing, see e.g., [22, 11, 12, 37, 15]. In all cases the sampling order depends only on the smoothness difference of the two continuous (semi–)norms involved.

3.2. Proof Sketch

A standard way to prove such sampling inequalities follows the lines of [24] and [37]. For some domain $\mathcal{D}$ star-shaped with respect to a ball, let $\left\{ a_j^{(\alpha)} : j = 1, \ldots, N \right\}$ be a polynomial reproduction of degree k with respect to a discrete set $X = \{x_1, \ldots, x_N\} \subset \mathcal{D}$, i.e.,

$$D^\alpha q(x) = \sum_{j=1}^N a_j^{(\alpha)}(x) q(x_j)$$

holds for every $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$, all $x \in \mathcal{D}$ and all $q \in \pi_k^d(\mathcal{D})$ where π_k^d denotes the space of all d-variate polynomials of degree not exceeding k. Then we have

$$
\begin{aligned}
&|D^\alpha u(x)| \\
\leq\ & |D^\alpha u(x) - D^\alpha p(x)| + |D^\alpha p(x)| \\
\leq\ & \|D^\alpha u - D^\alpha p\|_{L_\infty(\mathcal{D})} + \sum_{j=1}^N \left| a_j^{(\alpha)}(x) \right| |p(x_j)| \\
\leq\ & \|D^\alpha u - D^\alpha p\|_{L_\infty(\mathcal{D})} + \sum_{j=1}^N \left| a_j^{(\alpha)}(x) \right| \|S_X(p)\|_{\ell_\infty(\mathbb{R}^N)} \\
\leq\ & \|D^\alpha u - D^\alpha p\|_{L_\infty(\mathcal{D})} + \\
& + \sum_{j=1}^N \left| a_j^{(\alpha)}(x) \right| \left(\|u - p\|_{L_\infty(\mathcal{D})} + \|S_X u\|_{\ell_\infty(\mathbb{R}^N)} \right)
\end{aligned}
$$

for arbitrary $u \in W_p^k(\mathcal{D})$ and any polynomial $p \in \pi_k^d(\mathcal{D})$. Using a polynomial reproduction argument based on *norming sets* [21], the Lebesgue constant can be bounded by $\sum_{j=1}^N \left| a_j^{(\alpha)} \right| \leq 2$, if some moderate *over-sampling* is allowed which is controlled via a Markov inequality. As a polynomial approximation we choose the averaged Taylor polynomial of degree k ([14, Section 2]). This leads to a sampling inequality of the form

$$
\begin{aligned}
\|D^\alpha u\|_{L_\infty(\mathcal{D})} \quad \leq\ & \frac{C}{(k - |\alpha|)!} \delta_{\mathcal{D}}^{k - d/p} \left(\delta_{\mathcal{D}}^{-|\alpha|} + h^{-|\alpha|} \right) |u|_{W_p^k(\mathcal{D})} \\
& + 2h^{-|\alpha|} \|S_X u\|_{\ell_\infty(\mathbb{R}^N)}
\end{aligned}
$$

where $\delta_{\mathcal{D}}$ denotes the diameter of $\mathcal{D}$. To derive sampling inequalities on a Lipschitz domain Ω satisfying an interior cone condition, we cover Ω by domains $\mathcal{D}$ which are star-shaped with respect to a ball, satisfying $\delta_{\mathcal{D}} \approx h$ (see Duchon [18] for details on such coverings). Global estimates are obtained by summation or maximization over the local estimates.

3.3. Strong Sampling Inequalities: Infinite Smoothness

We now consider strong sampling inequalities for infinitely smooth functions in the sense of [27] where the sampling orders turn out to vary exponentially with the fill distance h. When applied to errors of discretization processes involving analytic functions, such sampling inequalities yield convergence results of exponential order, like those of spectral methods.

Following the above proof sketch, these exponential orders are achieved by appropriately coupling the parameter k, controlling the order of smoothness and the order of polynomial reproduction, to the fill distance h. The key point for relations between k and h is the existence of polynomial reproductions of order k on samplings of fill distance $\mathcal{O}(h)$. Following [27], we handle infinitely smooth functions on a domain Ω by normed linear function spaces $\mathcal{H}(\Omega)$ that can for some fixed $1 \leq p < \infty$ be continuously

embedded into every classical Sobolev space $W_p^k(\Omega)$. More precisely, for some $p \in [1, \infty)$ and all $k \in \mathbb{N}$ we assume that there are embedding operators $I_k^{(p)}$ and constants $E(k)$ such that

$$I_k^{(p)} : \mathcal{H}(\Omega) \to W_p^k(\Omega) \quad \text{with}$$
$$\left\| I_k^{(p)} \right\|_{\{\mathcal{H} \to W_p^k(\Omega)\}} \leq E(k) \quad \text{for all } k \in \mathbb{N}_0.$$

There are various examples of spaces with this property, e.g., Sobolev spaces of infinite order as they occur in the study of partial differential equations of infinite order [10], or reproducing kernel Hilbert spaces of Gaussians and inverse multiquadrics.

In the case of infinitely smooth functions, the sampling order is mainly determined by the asymptotic behaviour of the embedding constants $E(k)$ for $k \to \infty$. Typical examples are $E(k) = 1$ for $W_p^\infty(\Omega)$, or $E(k) \leq C^k k^{k/2}$ for the reproducing kernel Hilbert space of Gaussians [27]. A typical result is the following theorem [27]:

Theorem 3. *Suppose that $\Omega \subset \mathbb{R}^d$ is bounded, has a Lipschitz boundary, and satisfies an interior cone condition. If there are constants $0 < \epsilon \leq 1$ and $C_E > 0$ such that the embedding constants are bounded by $E(k) \leq C_E^k k^{(1-\epsilon)k}$ for all $k \in \mathbb{N}$, then, for all $1 \leq q \leq \infty$ there are constants C, C_1 and $h_0 > 0$ such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_0$ and sampling operators S_X from (5) the inequality*

$$\|D^\alpha u\|_{L_q(\Omega)} \leq e^{C \log(h)/\sqrt{h}} \|u\|_{\mathcal{H}(\Omega)} + C_1 h^{-|\alpha|} \|S_X u\|_{\ell_\infty(\mathbb{R}^N)}$$

holds for all $u \in \mathcal{H}(\Omega)$. Here, the constant C does not depend on the space dimension d.

Similar results for different classes of infinitely smooth functions are obtained in [39].

3.4. Weak Sampling Inequalities

Now we focus on *weak* sampling operators. We consider a set of arbitrary functionals

$$\Lambda = \{\lambda_1, \ldots, \lambda_N\} \subset \left(W_2^k(\Omega)\right)^*.$$

These functionals define a *weak sampling operator*

$$S_\Lambda := (\lambda_1, \ldots, \lambda_N)^T$$

and we expect a sampling inequality of the form

$$\|u\|_{L_2(\Omega)} \leq C h_\Lambda^k |u|_{W_2^k(\Omega)} + C \|S_\Lambda u\|_{\ell_\infty(\mathbb{R}^N)}.$$

Such an estimate can hold true only if the functionals Λ are a norming set [21] for the polynomials of degree less than k (see section 2.2). We will present two examples of such functionals which are of current research interest.

3.4.1. Weak Convolution-Type Data

Like in [28], we shall use *convolution-type data* of the form

$$\lambda_j(u) = \int_\Omega K(x - x_j) u(x)\, dx \qquad (15)$$

to build a sampling operator for weak data. Here $X = \{x_1, \ldots, x_N\} \subset \Omega$ is a discrete set of points, and $K : \mathbb{R}^d \to \mathbb{R}$ is called a *test kernel*. We shall consider only convolution-type data generated by translation invariant test kernels. These weak data are closely related to *Finite Volume Methods*. In the usual Finite Volume Method, one simply chooses $K(\cdot - x_j) \equiv 1$ on supp $K(\cdot - x_j)$. There is a theoretical consideration of this case in [35]. In [28], this situation is generalized to non-stationary convolution type data of the form (15), where the kernel K is of some fixed scale for all translates. In [25], the case of *stationary data* is considered, where the support of the test kernel $K(\cdot - \cdot)$ is scaled with the mesh-norm of X in Ω. This implies that $\lambda_j(f)$ is some weighted local mean of f. To be precise we shall impose the following conditions on K:

Definition 1. *A kernel $K : \mathbb{R}^d \to \mathbb{R}$ is called a stationary test kernel if it satisfies*

1. $\int_\Omega K(x - x_j)\, dx = 1$ *for all* $x_j \in X$,
2. $\mathrm{supp}(K(\cdot - x_j)) =: V_j \subset \Omega$,
3. $c_1 h_{X,\Omega} \leq \mathrm{diam}(V_j) \leq c_2 h_{X,\Omega}$, *and*
4. $\|K(\cdot - x_j)\|_{L_p(\Omega)} \leq C h_{X,\Omega}^{-d/q}, \quad \frac{1}{p} + \frac{1}{q} = 1$.

Under these conditions there is the following theorem from [25].

Theorem 4. *For weak sampling functionals (15) using a stationary test kernel, there is a sampling inequality*

$$\|u\|_{L_2(\Omega)} \leq C\left(h_{X,\Omega}^k \|u\|_{W_2^k(\Omega)} + \|S_\Lambda(u)\|_{\ell_\infty(\mathbb{R}^N)}\right) \qquad (16)$$

for all $u \in W_2^k(\Omega)$, $k \geq 0$ and all sets X of centers with sufficiently small fill distance $h_{X,\Omega}$.

Such a sampling inequality will be very useful in the error analysis of numerical methods for weak solutions of partial differential equations, since it yields error estimates for functions in $W_2^1(\Omega)$. These estimates are not covered by the well–established theory since $W_2^1(\Omega)$ is not continuously embedded in $C(\Omega)$ for $\Omega \subset \mathbb{R}^d$ with $d \geq 2$.

3.4.2. Galerkin Methods

Another kind of weak data, which arises naturally in the study of partial differential equations, is *Galerkin data*. Namely, we consider a partial differential equation in its weak formulation

$$\text{find } u \in H \; : \; a\,(u,v) = F\,(v) \quad \text{for all } v \in X, \tag{17}$$

where $H = W_2^\tau\,(\Omega)$ is typically a Sobolev space and $a : H \times H \to \mathbb{R}$ is a bilinear form while $F \in H^*$ is a linear functional. Usually, the solution to (17) lies actually in a subspace $W_2^\sigma\,(\Omega) \subset W = W_2^\tau\,(\Omega)$ of order $\sigma > \tau$. This difference in smoothness leads to some sampling order. To solve the problem (17) approximatively we use a Ritz-Galerkin approach [14, (2.5.7)] and consider the finite dimensional problem

$$\text{find } u \in V_{K,X} \; : \; a\,(u,v) = F\,(v) \; \text{ for all } v \in V_{K,X},$$

where

$$V_{K,X} := \text{span } \{K\,(\cdot - x_j) \; : \; x_j \in X\}$$

with a translation invariant kernel K. To simplify the constants in our sampling inequality, it is convenient to consider an orthonormal basis $\{\phi_j\}_{j=1,\dots,N}$ of $V_{K,X}$ with respect to the bilinear form $a\,(\cdot,\cdot)$. We define a sampling operator $S_\Lambda := (\lambda_1, \dots, \lambda_N)^T$ by

$$\lambda_j\,(u) := a\,(u, \phi_j)\,.$$

If there the unique solution u^* to (17) satisfies $u^* \in W_2^k$ (see, e.g., [14] for a formulation of this condition in terms of a and F), then there is a sampling inequality of the form (11) with $\sigma = k - \tau$ (see [25, Thm. 8.3.1] for details).

§4. Sampling and Stability in Reconstruction Processes

The previous sections dealt with sampling and stability inequalities in general and in particular, but now we turn to *applications* of both. These can be described as general *reconstruction problems*, where one tries to recover a function f from a weak or strong data provided by a sampling $S_\Lambda\,(f)$.

4.1. Testing Trial Functions

Starting from a set

$$\Lambda_E = \{\lambda_1, \dots, \lambda_{N_E}\} \subset \mathcal{F}^*$$

of linear functionals, we define a *sampling operator* via

$$S_{\Lambda_E} := (\lambda_1, \ldots, \lambda_{N_E})^T.$$

Then we consider given data $S_{\Lambda_E}(f)$ generated by application of the sampling operator S_{Λ_E} to an unknown function f, and we *test* certain *trial functions* u_r to see if they reproduce the data, i.e., if

$$S_{\Lambda_E}(u_r) = S_{\Lambda_E}(f) \tag{18}$$

holds. Thus *sampling* here is the same as *testing*. This viewpoint is borrowed from Petrov–Galerkin methods, where an approximate solution is taken from a space of *trial functions*, and it has to satisfy a number of *test equations*. There is a lot of literature in this field, the following collection is by far not complete, but should give an overview [1, 2, 3, 4]. In the Finite Element Method, these test equations are *weak*, since they are integrals against test functions, while in collocation methods the test equations are *strong*, i.e., they are evaluations of functions or their derivatives at discrete points. Evaluating test data on trial functions is nothing else than *sampling* them in the sense of this survey.

Consequently, we shall carefully distinguish between the *test* and the *trial* side. The *test* side consists of the sampling operator S_{Λ_E} based on the given functionals Λ_E. The *trial* side consists of a finite dimensional *trial space* $\mathcal{R}$ which is used to generate an approximate solution $u_r \in \mathcal{R}$ to the problem. Typical examples for trial spaces are linear hulls of translates of a kernel or finite element spaces. Note that we use letters like $\mathcal{R}$, r etc. for the t**R**ial side and E, ϵ etc. for the t**E**st side. In applications to PDE solving, sampling and testing involves a large variety of functionals. For instance, in strong collocation methods for a Poisson problem

$$
\begin{array}{rcll}
-\Delta u & = & g & \text{in } \Omega \\
u & = & \varphi & \text{on } \partial\Omega
\end{array}
$$

there will be functionals of the forms

$$
\begin{array}{rcll}
\lambda_j(u_r) & := & -\Delta u_r(x_j), & x_j \text{ in } \Omega, \quad 1 \le j \le N_\Omega \\
\lambda_j(u_r) & := & u_r(x_j), & x_j \text{ on } \partial\Omega, \quad N_\Omega < j \le N_E,
\end{array}
$$

and we assume that there is some f that provides sampled data $\lambda_j(f)$ to yield the test equations

$$
\begin{array}{rclclcll}
\lambda_j(u_r) & = & -\Delta u_r(x_j) & = & -\Delta f(x_j) & = & g(x_j), & 1 \le j \le N_\Omega \\
\lambda_j(u_r) & = & u_r(x_j) & = & f(x_j) & = & \varphi(x_j), & N_\Omega < j \le N_E
\end{array}
$$

which are at least approximately solved for a trial function u_r. Weak methods will replace point evaluations by local integrals against test functions,

but our common viewpoint is that trial functions u_r are tested by sampling them and comparing the sampled data $\lambda_j (u_r)$ to the given sampled data $\lambda_j (f)$ of an unknown exact solution f.

Error estimates will take the form of upper bounds of $\|u_r - f\|$ in some norm. But if $dist (f, \mathcal{R})$ is large, there is no testing strategy that can lead to small errors. Thus users must care first for a trial space $\mathcal{R}$ which allows good approximations u_r to the function f providing the test data. This fact is often overlooked, and it is independent of how testing is done. The job of testing is to enable the user to select one good approximation $u_r \in \mathcal{R}$ to f in an efficient and stable way. Thus the *error* of the reconstruction process depends crucially on the *trial* side, while the *test* side cares for the *stability* of the reconstruction.

When testing trial functions, *interpolation* requires the trial functions to reproduce the test data exactly, whereas *approximation* allows small errors in data reproduction. While in some applications interpolation is required, others, in particular those involving errors or noise in the given data, prefer approximation methods. Sometimes certain parameters are used to control the error in data reproduction, e.g., in machine learning or spline smoothing, and these parameters are closely related to regularization. Examples will follow later, after we have described our favourite trial spaces.

4.2. Kernel–Based Trial Spaces

Throughout the rest of this survey, we shall mainly restrict ourselves to kernel-based methods, i.e., to "meshless" trial spaces spanned by translates of kernels. In their simplest form, they define a trial space $\mathcal{R}$ generated by translates of a single *trial kernel* $K (\cdot, \cdot)$ via

$$\mathcal{R} := \operatorname{span} \{ K (x_j, \cdot) \ : x_j \in X_R \},$$

where the set $X_R := \{x_1, \ldots, x_{N_R}\} \subset \Omega$ is called the set of *trial points*. To avoid complications, our trial kernels $K \ : \ \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ will always be symmetric, continuous and positive definite on $\mathbb{R}^d$, though we shall often restrict them to a domain $\Omega \subseteq \mathbb{R}^d$. They are reproducing kernels in a *native* Hilbert space [36] $\mathcal{N}_K(\Omega)$ of functions on $\Omega \subseteq \mathbb{R}^d$ in the sense

$$f (x) = (f, K (x, \cdot))_{\mathcal{N}_K} \ \text{for all } f \in \mathcal{N}_K(\Omega), \ x \in \Omega \subseteq \mathbb{R}^d,$$

and $\mathcal{N}_K (\Omega)$ is continuously embedded in $C (\Omega)$. Typical examples are Gaussians $K (x, y) = \exp \left(-\|x - y\|_2^2 \right)$ or Sobolev–Matérn kernels

$$K_{\tau - d/2} \left(\|x - y\|_2 \right) \|x - y\|_2^{\tau - d/2}$$

for $\tau > d/2$ with the Bessel functions K_ν of second kind. In the latter case one has $\mathcal{N}_K (\Omega) = W_2^\tau (\Omega)$, and certain compactly supported *Wendland*

kernels are reproducing in Hilbert spaces which are norm–equivalent to Sobolev spaces. See [36] for details. To arrive at useful sampling orders, we shall assume a continuous embedding

$$\mathcal{N}_K(\Omega) \subseteq W_2^\tau(\Omega) \tag{19}$$

for some $\tau > 0$. Concerning the dependence of native Hilbert spaces $\mathcal{N}_K(\Omega)$ on the domain Ω and related extension/restriction mappings, we again have to refer to [36] for details. For a detailed overview on the use of kernels, in particular in partial differential equations, see the recent review [30] and the references therein.

By picking trial kernels of limited smoothness, we confine ourselves in this section to the case of finitely smooth trial spaces. This is just for the sake of simplicity. Most of the issues can be carried over to the infinitely smooth case using analytic kernels like the Gaussian.

Kernel–based trial spaces allow to align the trial side formally with the test side. This is done by a second set

$$M_R = \{\mu_1, \dots, \mu_{N_R}\} \subset \mathcal{N}_K^*(\Omega)$$

of linear functionals and a second *sampling operator* for trial purposes via

$$S_{M_R} := (\mu_1, \dots, \mu_{N_R})^T$$

to define a generalized kernel–based trial space

$$\mathcal{R}_{M_R} := \operatorname{span}\left\{\mu_j^x K(x, \cdot) \ : \mu_j \in M_R\right\} \subset \mathcal{N}_K(\Omega) \tag{20}$$

where we denote the action of μ_j on the variable x by μ_j^x. The two sampling operators S_{M_R} and S_{Λ_E} will allow us to use sampling inequalities on both the trial and the test side.

Introducing coefficient vectors $\alpha = (\alpha_1, \dots, \alpha_{N_R})^T \in \mathbb{R}^{N_R}$ we can write the trial functions from $\mathcal{R}_{M_R}$ as

$$u_\alpha(\cdot) := \sum_{j=1}^{N_R} \alpha_j \mu_j^x K(x, \cdot) = \alpha^T S_{M_R}^x K(x, \cdot). \tag{21}$$

4.3. Symmetric Interpolation Methods

In *symmetric methods* the test and trial sides are the same, i.e., we set $S_{M_R} = S_{\Lambda_E}$ and $N_R = N_E$, allowing to skip the R and E indices for this subsection and to use $\Lambda = \{\lambda_1, \dots, \lambda_N\} \subset \mathcal{N}_K(\Omega)^*$ throughout. Each trial function will be of the form (21), and the interpolation system (18) takes the form of a linear equation for the coefficient vector α

$$S_\Lambda(u_\alpha) = A_{\Lambda,\Lambda}\alpha = S_\Lambda(f), \tag{22}$$

where the *kernel* matrix

$$A_{\Lambda,\Lambda} := \left(\lambda_i^x \lambda_j^y K(x,y)\right)_{1 \le i,j \le N} \tag{23}$$

is symmetric and positive definite if the functionals in $\Lambda \subset \mathcal{N}_K(\Omega)^*$ are linearly independent and if the kernel K is positive definite. Under these assumptions, the above system is uniquely solvable [38] for data given by arbitrary functions $f \in \mathcal{N}_K(\Omega)$.

We now sketch how sampling inequalities can be used for a fully general error analysis of symmetric interpolation. We point out that sampling inequalities can also be used for non-interpolatory and unsymmetric recovery processes, as will be shown in later sections.

Assume that the sampling operator S_Λ allows a sampling inequality of the form (4) on a space $\mathcal{F}$ into which $\mathcal{N}_K(\Omega)$ can be continuously embedded. Then we have

$$\|f - u^*(f)\|_W \le Ch_\Lambda^\sigma |f - u^*(f)|_{\mathcal{F}} \quad \text{for all } f \in \mathcal{N}_K(\Omega)$$

if we denote the solution of the interpolation system by $u^*(f)$. Using the embedding $\mathcal{N}_K(\Omega) \subset \mathfrak{S}$ and the standard minimum–norm property [36, 38]

$$\|u^*(f)\|_{\mathcal{N}_K(\Omega)} \le \|f\|_{\mathcal{N}_K(\Omega)}$$

of solutions of systems like (22) this implies the error bound

$$\|f - u^*(f)\|_W \le Ch_\Lambda^\sigma \|f\|_{\mathcal{N}_K(\Omega)} \quad \text{for all } f \in \mathcal{N}_K.$$

We shall give a few examples.

4.3.1. Symmetric Strong Interpolation

The simplest case takes a finite subset X of Ω with fill distance (6) and assumes (19). Then we can use the above argument to apply Theorem 1 and get an error bound

$$\|f - u^*(f)\|_{L_q(\Omega)} \le Ch^{\tau - d\left(\frac{1}{2} - \frac{1}{q}\right)_+} \|f\|_{\mathcal{N}_K(\Omega)} \tag{24}$$

for all $f \in \mathcal{N}_K$ and all sets X with sufficiently small fill distance h. This reproduces the well known error bounds (see [36]). If some other functionals are added, such a bound will still hold.

4.3.2. Symmetric Weak Interpolation

In the case of weak convolution-type data based on a set $X = \{x_1, \ldots, x_N\}$ as defined in (15), i.e.,

$$\lambda_j(u) = \int_\Omega \mathfrak{E}(x - x_j) u(x)\, dx \quad j = 1, \ldots, N$$

with a stationary test kernel $\mathfrak{E}$ as in Definition 1 the kernel matrix $A_{\Lambda,\Lambda}$ as defined in (23) has entries

$$(A_{\Lambda,\Lambda})_{i,j} := \int_\Omega \int_\Omega \mathfrak{E}\left(x - x_i\right) \mathfrak{E}\left(y - x_j\right) K\left(x, y\right) dx dy$$

where the kernel K is symmetric, positive definite and satisfies (19). Then the sampling inequality of the form (16) yields as a direct consequence

Theorem 5. *Under the above assumptions, recovery of functions f by symmetric weak interpolation leads to interpolants $u^*\left(f\right)$ with error bounds*

$$\|f - u^*\left(f\right)\|_{L_2(\Omega)} \leq C h_{X,\Omega}^\tau \|f\|_{\mathcal{N}_K(\Omega)} \tag{25}$$

for all $f \in \mathcal{N}_K\left(\Omega\right)$ and all discrete sets X with sufficiently small fill distance $h_{X,\Omega}$.

4.4. Symmetric Non-Interpolatory Methods

Although interpolation processes lead to good approximation properties, they have some drawbacks, e.g., the condition of the system (22) is dominated by the separation distance q_X of (7) which can be considerably smaller than the fill distance. In particular, if an additional point comes close to a point of X, the system condition deteriorates dramatically, but omission of the new point would do no harm and can be viewed as a regularization of the augmented system. This indicates that regularized methods are important even in the case of noise-free data. We shall deal with these now.

4.4.1. Approximation as Regularized Interpolation

Here, we outline how sampling inequalities can be used to derive worst-case convergence rates for regularized reconstruction processes. We shall concentrate on regularization methods that avoid exact solving of the system (22).

Besides improving condition numbers, most regularization methods have several advantages, e.g., regularization is closely related to *sparse approximation* [19]. The crucial point for the analysis of all regularized reconstruction processes Π^ν, where ν is a regularization parameter, is to show the following two properties

$$\begin{aligned} \|\Pi^\nu\left(f\right)\|_{W_2^\tau(\Omega)} &\leq& \|f\|_{W_2^\tau(\Omega)} \quad \text{and} \\ \max_{1 \leq j \leq N} |\lambda_j\left(f - \Pi^\nu f\right)| &\leq& g\left(\nu, f\right) \|f\|_{W_2^\tau(\Omega)} \end{aligned}$$

where the function $g\left(\nu, f\right)$ determines the approximation quality of Π^ν. These properties can be seen as *stability* and *consistency*. We give two examples where these properties are successfully coupled with sampling inequalities.

4.4.2. Spline Smoothing

We shall focus on the well-known case of *spline smoothing* or ℓ_2-*spline-regression*. A more detailed overview can be found in [34] and [36]. For a given $f \in W_2^\tau(\Omega)$ and the various functionals λ_j from the previous sections we can formulate the smoothed optimal recovery problem

$$\min_{u \in W_2^\tau(\Omega)} \left(\sum_{j=1}^N |\lambda_j(u - f)|^2 + \nu \|u\|_{W_2^\tau(\Omega)}^2 \right), \tag{26}$$

where $\nu \geq 0$ is called the *smoothing parameter*. The case of strong data $\lambda_j = \delta_{x_j}$ is discussed in [37], whereas the case of weak convolution-type data is dealt with in [25]. For a detailed discussion of the smoothing parameter see [34]. We simply note that the special case $\nu = 0$ corresponds to finding a generalized interpolant, i.e., a function $u^{(0)}(f) \in W_2^\tau(\Omega)$ that satisfies the generalized interpolation conditions (18). It is well known [36] that there always exists a solution $u^{(\nu)}(f)$ to this relaxed interpolation problem (26) in the linear space (20). The coefficients $\alpha \in \mathbb{R}^N$ with respect to the basis $\{\lambda_j^x K(\cdot, x)\}$ can be found by solving the linear system

$$(A_{\Lambda,\Lambda} + \nu \mathrm{Id})\alpha = f_\Lambda,$$

with the kernel matrix $A_{\Lambda,\Lambda}$ from (23) and

$$f_\Lambda = S_\Lambda(f) = (\lambda_1(f), \ldots, \lambda_N(f))^T.$$

As elaborated in [37] for strong data, we have the following two inequalities

$$\left\| u^{(\nu)}(f) \right\|_{W_2^\tau(\Omega)} \leq \left\| u^{(\nu)}(f) \right\|_{W_2^\tau(\Omega)} \quad \leq \quad \|f\|_{W_2^\tau(\Omega)},$$
$$\left\| S_\Lambda \left(f - s^{(\nu)}(f) \right) \right\|_{\ell_\infty(\mathbb{R}^N)} \quad \leq \quad \sqrt{\nu} \|f\|_{W_2^\tau(\Omega)}.$$

Inserting into the sampling inequality (1) yields the bound

$$\left\| f - s_f^{(\nu)} \right\|_{L_2(\Omega)} \leq C \left(h^\tau + \sqrt{\nu} \right) \|f\|_{W_2^\tau(\Omega)}.$$

This inequality suggests an *a priori* choice of the smoothing parameter as $\nu \leq h^{2\tau}$, which leads to the optimal order [29] achievable by interpolation, but now using a linear system with a condition independent of the separation distance q_X.

4.4.3. Kernel-Based Learning

There is a close link between the theories of kernel-based approximation, spline smoothing and machine learning. Since there is a broad collection of literature on this topic (see, e.g., [31, 16, 30, 5, 6, 7, 8, 9] and the references

therein), we shall only briefly sketch the viewpoint of learning theory. In this section, we deal only with *strong* recovery or approximation problems. In order to recover a continuous function f from a strong sample $S_X(f) = (f(x_1), \ldots, f(x_N))^T$, the most common choice for an approximant in kernel-based approximation theory u_f is an interpolant, i.e., $S_X(u_f) = S_X(f)$. This method obviously makes the best use of the sample $S_X(f) \in \mathbb{R}^N$. But there are also some drawbacks of classical interpolation. On the one hand, the reconstruction is very unstable if we consider $S_X(f)$ to be spoiled by noise. On the other hand, there are also numerical disadvantages, namely the computation of the interpolant may be ill-conditioned. Furthermore, if $u_f(\cdot) = \sum_{j=1}^{N} \alpha_j K(\cdot, x_j)$ denotes the interpolant, there will be many non-zero coefficients, i.e., this method is non-sparse.

One way out of these problems is provided by learning theory. Here, the reconstruction problem is an example of *supervised regression*, because the real values $S_X(f)$ are generated by an unknown, but fixed function f. Instead of $S_X(f)$ we may consider more generally a vector of possibly disturbed values $y \approx S_X(f)$. One typically relaxes the interpolation condition by using a more general similarity measure, e.g., by using a *loss function*. A typical example is Vapnik's ϵ-intensive loss function [33]

$$|f(x) - y|_\epsilon = \left\{ \begin{array}{cc} 0 & if \; |f(x) - y| \leq \epsilon \\ |f(x) - y| - \epsilon & if \; |f(x) - y| > \epsilon \end{array} \right. ,$$

which allows deviations up to a positive parameter $\epsilon > 0$. A popular reconstruction technique called ν–Support–Vector–Regression (ν–SVR, [32]) in a Hilbert space $H(\Omega)$ of functions $\Omega \to \mathbb{R}$ is then obtained as solution to the optimization problem

$$\min_{\substack{u \in H(\Omega) \\ \epsilon \in \mathbb{R}^+}} \frac{1}{N} \sum_{j=1}^{N} |u(x_j) - y_j|_\epsilon + \epsilon \nu + \lambda \|u\|_{H(\Omega)}^2 \tag{27}$$

with an *a priori* chosen parameter λ. In this section we shall focus on the ν–SVR, but everything works as well for similar regression techniques such as the ϵ-SVR [27, 26]. The ν–SVR possesses a solution (u_y, ϵ^*) [23], and if $H(\Omega)$ is equivalent to the native Hilbert space of a symmetric positive definite kernel K, there is a solution of the form $u_y(\cdot) = \sum_{j=1}^{N} \alpha_j K(\cdot, x_j)$ which can be computed by solving a finite dimensional quadratic optimization problem for the coefficient vector α [31]. As elaborated in [26], sampling inequalities can be used to provide a worst-case error analysis, even in the case of noisy data. Such bounds require no knowledge about the underlying noise model. As outlined above, we assume that the data $y = y_f \in \mathbb{R}^N$ comes from the strong sample $S_X(f)$ from some function $f \in W_2^\tau(\Omega) \cong \mathcal{N}_K(\Omega)$ on a set $X = \{x_1, \ldots, x_N\} \subset \Omega$, but we allow the

given data to be corrupted by some additive error $r \in \mathbb{R}^N$, i.e.,

$$S_X(f) = y + r. \tag{28}$$

If we solve the optimization problem (27) with data $y = S_X(f) - r$, we obtain an approximant u_y to the generating function f. The following theorem provides an upper bound for the approximation error of this process.

Theorem 6. *Suppose some bounded domain $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary, and $\lfloor \tau - 1 \rfloor > d/2$. Then there are constants $C, h_0 > 0$, depending only on Ω, d, q and τ with the following property. For all function $f \in W_2^\tau(\Omega)$, all sets $X = \{x_1, \ldots, x_N\} \subset \Omega$ with fill distance $h \leq h_0$ and all errors $r \in \mathbb{R}^N$, a solution (u_y, ϵ^*) of (27) with data y being related to the samples of f via (28), satisfies for any $\delta > 0$*

$$\|f - u_y\|_{L_\infty(\Omega)} \leq C \left(h^{\tau - \frac{d}{2}} \left(\|f\|_{W_2^\tau(\Omega)} + \sqrt{\frac{1}{N\lambda} \sum_{j=1}^N |r_j|_\delta + \frac{\nu\delta}{\lambda} + \|f\|_{W_2^\tau(\Omega)}^2} \right) \right.$$
$$\left. + \sum_{j=1}^N |r_j|_\delta + \nu N\delta + \epsilon^*(1 - N\nu) + N\lambda \|f\|_{W_2^\tau(\Omega)}^2 \right).$$

The bound in Theorem 6 involves a positive parameter δ, which we are free to choose. The optimal choice of δ obviously depends on the problem parameters. Again, the error estimate suggests some choice of the problem parameters. If we assume that the error does not exceed the data itself, i.e., $\|r\|_{\ell_\infty(\mathbb{R}^N)} \leq \delta \leq \|f\|_{W_2^\tau(\Omega)}$, we can choose the parameters λ, ν and ϵ (see [26] for details) such that

$$\|f - u_y\|_{L_\infty(\Omega)} \leq C \left(h^{\tau - d/2} \|f\|_{W_2^\tau(\Omega)} + \delta \right).$$

This shows that the solution of the ν–SVR leads to optimal sampling orders [29] in Sobolev spaces with respect to the fill distance h. These optimal rates are also attained by classical interpolation in the native Hilbert space [36]. The ν–SVR, however, allows for much more flexibility and less complicated solutions.

4.5. Unsymmetric Methods

We now go back to Section 4.4.1 and deal with unsymmetric recovery methods where the sampling operator S_{Λ_S} used for testing is different from the sampling operator S_{M_R} defining the trial space. We simplify the trial space to let the sampling operator

$$S_{M_R} = \left(\delta_{x_1}, \ldots, \delta_{x_{N_R}} \right)^T$$

consist of point evaluations on a set of points $X_R := \{x_1, \ldots, x_{N_R}\} \subset \Omega$ only, and thus we use the trial space

$$\mathcal{R} := \text{span} \{K(x_j, \cdot) : x_j \in X_R\}.$$

We want to recover an unknown function $f \in \mathcal{N}_K(\Omega) \subseteq W_2^\tau(\Omega)$ from its given weak or strong test data

$$S_{\Lambda_E}(f) := (\lambda_1(f), \ldots, \lambda_{N_E}(f))^T$$

using trial functions from $\mathcal{R}$. That is, we seek a function $u_r(f) \in \mathcal{R}$ satisfying

1. $S_{\Lambda_E}(u_r(f)) \approx S_{\Lambda_E}(f)$
2. $\|u_r(f) - f\|_{W_2^m(\Omega)} \leq \epsilon(r, f, \tau, m)$ small, for some $m < \tau$.

Typically, one has $\epsilon(r, f, \tau, m) = h_R^{\tau - m} \|f\|_{W_2^\tau(\Omega)}$. Consider the interpolation system

$$S_{\Lambda_E}(u_r(f)) = S_{\Lambda_E}(f)$$

as in (18). Unfortunately, this system is unsymmetric and hence we cannot simply solve a linear system as in the case of symmetric methods. But we use the observation from Section 2.3 that under the condition (13) the linear system has full rank. Hence we can apply any numerical method solving the usually overdetermined system (18) to some accuracy by *residual minimization*. We denote by $u^*(f)$ the best approximant from $\mathcal{R}$ to f which is an interpolant to the unknown data $R_{M_R}(f)$ and define $u_r(f)$ by the following property

$$\|S_{\Lambda_E}(f) - S_{\Lambda_E}(u_r(f))\|_{\mathbb{R}^{N_E}} = \inf_{u_r \in \mathcal{R}} \|S_{\Lambda_E}(f) - S_{\Lambda_E}(u_r)\|_{\mathbb{R}^{N_E}}.$$

Then we have

$$\begin{aligned} \|S_{\Lambda_E}(f) - S_{\Lambda_E}(u_r(f))\|_{\mathbb{R}^{N_E}} &= \inf_{u_r \in \mathcal{R}} \|S_{\Lambda_E}(f) - S_{\Lambda_E}(u_r)\|_{\mathbb{R}^{N_E}} \\ &\leq \|S_{\Lambda_E}(f) - S_{\Lambda_E}(u^*(f))\|_{\mathbb{R}^{N_E}} \\ &\leq \epsilon(r, f, \tau, m)\|S_{\Lambda_E}\|, \end{aligned}$$
(29)

where $\|S_{\Lambda_E}\|$ denotes the norm of the test sampling operator as a mapping between $W_2^m(\Omega)$ and $\mathbb{R}^{N_E}$. Note that we used a sampling inequality to bound the interpolation error. This gives the first property. For the second we find

$$\|f - u_r(f)\|_{W_2^m(\Omega)} \leq \|f - u^*(f)\|_{W_2^m(\Omega)} + \|u_r(f) - u^*(f)\|_{W_2^m(\Omega)}.$$

In order to bound the second term on the right-hand side, we apply a stability inequality of the form (10) to the function $u_r(f) - u^*(f) \in \mathcal{R}$

and get

$$\|u_r\left(f\right) - u^*\left(f\right)\|_{W_2^m(\Omega)}$$
$$\leq \quad C_{stab}\left(m, h_E, \mathcal{R}\right) \|S_{\Lambda_E}\left(u_r\left(f\right)\right) - S_{\Lambda_E}\left(u^*\left(f\right)\right)\|_{\mathbb{R}^{N_S}}$$
$$\leq \quad \epsilon(r, f, \tau, m) C_{stab}\left(m, h_E, \mathcal{R}\right) \|S_{\Lambda_E}\|.$$

Now everything combines into

$$\|f - u_r\left(f\right)\|_{W_2^m(\Omega)} \leq C\epsilon(r, f, \tau, m)\left(1 + C_{stab}\left(m, h_E, \mathcal{R}\right) \|S_{\Lambda_E}\|\right)$$

which reduces under the assumption $\epsilon(r, f, \tau, m) = h_R^{\tau - m}\|f\|_{W_2^\tau(\Omega)}$ to the usual form

$$\|f - u_r\left(f\right)\|_{W_2^m(\Omega)} \leq Ch_R^{\tau - m}\left(1 + C_{stab}\left(m, h_E, \mathcal{R}\right) \|S_{\Lambda_E}\|\right)\|f\|_{W_2^\tau(\Omega)}.$$

4.6. Unsymmetric Weak Recovery

The paper [28] deals with unsymmetric recovery of functions $f \in L_2(\Omega)$ from weak data obtained via nonstationary convolution. Under suitable conditions on the trial and test kernels, the numerical solution $u_r(f)$ has an error bound of the form

$$\|f - u_r(f)\|_{W_2^{-2\tau}} \leq Ch_R^{2\tau}\|f\|_{L_2(\Omega)} \text{ for all } f \in L_2(\Omega)$$

provided that both kernels have at least smoothness τ in the sense of section 4.2. To show how weak stationary sampling can be put into the framework of the previous section, we follow [25] and combine the functionals from Section 3.4.1 with the ideas from the last section and from [28] to derive L_2–error estimates. For convenience we briefly recall the definition of *weak stationary convolution-type data*

$$\lambda_j\left(u\right) = \int_\Omega \mathfrak{E}\left(x - x_j\right) u\left(x\right) dx$$

with a stationary test kernel $\mathfrak{E}$ as characterized in Definition 1. As outlined in Section 4.5 we consider the following problem. An unknown function $f \in W_2^\tau\left(\Omega\right)$ has to be recovered approximately from its data $\left(\lambda_1\left(f\right), \ldots, \lambda_N\left(f\right)\right)^T$. We know that there is a good but unknown approximation $u^*\left(f\right)$ from the trial space $\mathcal{R} = \mathrm{span}\left\{K\left(x_j, \cdot\right) : x_j \in X_R\right\}$ to the function $f \in W_2^\tau\left(\Omega\right)$. Under certain conditions [25], there is an error estimate of the form

$$\|f - u^*(f)\|_{L_2(\Omega)} \leq h_R^\tau \|f\|_{W_2^\tau(\Omega)}, \tag{30}$$

showing the expected approximation order. The unsymmetric overdetermined system (18) takes the form

$$\lambda_j\left(f - u_r\right) \quad = \quad 0 \text{ for all } 1 \leq j \leq N_E.$$

As shown in (29) we can find a function $u_r(f) \in \mathcal{R}$ which solves this system up to some accuracy, i.e.,

$$|\lambda_j(f - u_r(f))| \leq Ch_R^\tau \|S_{\Lambda_E}\| \|f\|_{W_2^\tau(\Omega)} \text{ for } 1 \leq j \leq N_E.$$

As pointed out in section 4.5 and in (10) we assume an inverse inequality of the form

$$\|u\|_{W_2^\tau(\Omega)} \leq C\gamma(X_R) \|u\|_{L_2(\Omega)} \quad \text{for all } u \in \mathcal{R}.$$

Unfortunately, the value of $\gamma(X_R)$ is in general not known. There is a result in this direction in [25], namely if the trial kernel K is a radial basis function with algebraic smoothness $\tau > d/2$ and if X_R is separated from the boundary by q_R, then $\gamma(X_R) \approx q_R^{-\tau}$. Here, q_R denotes the separation distance of X_R. However, we can always make sure that the fill distance of the test data is small enough to stabilize the reconstruction, i.e., we shall assume a coupling of the discretizations as in (13),

$$\gamma(X_R) \, h_E^\tau \leq \frac{1}{2C} \quad \text{with } C > 1. \tag{31}$$

Combining everything yields the following result:

Theorem 7. *We denote by $u_r(f) \in \mathcal{R}$ the approximate solution of the system (18). Then we have an error bound of the form*

$$\|f - u_r(f)\|_{L_2(\Omega)} \leq C \left(2h_E^\tau + \frac{1}{C_1} h_R^\tau + h_R^\tau \|S_{\Lambda_E}\| \right) \|f\|_{W_2^\tau(\Omega)}$$

for sufficiently fine test discretizations.

In contrast to the results of [28], this error analysis does not assume f to be known on a slightly larger domain $\tilde{\Omega}$. Furthermore, the result holds for the L_2-norm, not for negative order Sobolev norms as in [28]. Unfortunately, the norm of the sampling operator shows up in the final estimate. This norm $\|S_{\Lambda_E}\| \sim h_S^{-d/2}$ grows and hence prevents optimal rates in the error estimate. Furthermore one needs to sassume $\tau > d/2$, which excludes fully weak problems. A promising way out seems to be to drop the inadequate ℓ_∞ norm in favor of $h^{d/2}\|S_{\Lambda_E}u\|_{\ell_2(\mathbb{R}^N)}$ like in Madych's sampling inequality (14). But this is work in progress.

§5. Outlook

Research on general sampling and stability inequalities is just in its beginnings, but since they are of central importance to Numerical Analysis, they deserve a larger audience via this survey. There are at least two obvious directions for future research. On the one hand-side, there are many

more possible applications for which sampling inequalities involving different functionals are required. This might include sampling of parametric curves and surfaces. On the other hand, there are some technical issues for improving the practical applicability. In every theorem, there are several generic constants involved for which one would need better bounds. This interacts with the stability estimates and the coupling of trial and test discretizations. In particular, to provide realistic upper bounds on fill distances on the test side, various generic constants need thorough investigation.

§6. Acknowledgement

We would like to thank the referees and the editor for several useful comments.

References

1. Y. Zhang Convergence of meshless Petrov-Galerkin method using radial basis functions *Applied Mathematics and Computation*, 183(1):307-321, 2006

2. I. Raju, D. Phillips and T. Krishnamurthy A radial basis function approach in the meshless local Petrov-Galerkin method for Euler-Bernoulli beam problems *Computational Mechanics*, 34(6):464-474, 2004

3. S. Atluri and S. Shen The Meshless Local Petrov-Galerkin (MLPG) Method: A Simple & Less-costly Alternative to the Finite Element and Boundary Element Methods *CMES* 3(1):11-51, 2002

4. S. Atluri and S. Shen The meshless local Petrov-Galerkin (MLPG) method Tech Science Press, Encino, 2002

5. T. Poggio and S. Smale The Mathematics of Learning: Dealing with Data *Notices of the AMS*, 50(5):537-544, 2003

6. F. Cucker and S. Smale On the mathematical foundations of learning *Bulletin of the AMS*, 39(1):1-49, 2001

7. S. Smale and D. Zhou Estimating the approximation error in learning theory *Analysis and Applications*,1(1):1-25, 2003

8. A. Caponnetto and E. DeVito Optimal Rates for the Regularized Least-Squares Algorithm *Found.Comput.Math.*, 331-368(2007)

9. A. Caponnetto and E. DeVito Learning and approximation by Gaussians on Riemannian manifolds *Adv Comput Math*,29:291-310,2008

10. A. Agadzhanov. Functional properties of Sobolev spaces of infinite order. *Soviet. Math. Dokl.*, 38(1):88–92, 1989.

11. R. Arcangéli, M. López de Silanes, and J. Torrens. An extension of a bound for functions in Sobolev spaces, with applications to (m, s)-spline interpolation and smoothing. *Numer. Math.*, 107:181–211, 2007.

12. R. Arcangéli, M. López de Silanes, and J. Torrens. Estimates for functions in Sobolev spaces defined on unbounded domains. to appear in Journal of Approximation Theory, 2008.

13. P. Borwein and T. Erdelyi. *Polynomials and Polynomial Inequalities*. Springer, New York, 1995.

14. S.C. Brenner and L.R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, 1994.

15. A. Corrigan and J. Wallin and T. Wanner. A sampling inequality for fractional order Sobolev semi-norms using arbitrary order data. preprint 2008.

16. F. Cucker and D.-X. Zhou. *Learning Theory: An Approximation Theory Viewpoint*. Cambridge University Press, 2007.

17. St. De Marchi and R. Schaback. Stability of kernel-based interpolation. to appear in Advances in Computational Mathematics, 2008.

18. J. Duchon. Sur l'erreur d'interpolation des fonctions de plusieurs variables pas les D^m–splines. *Rev. Française Automat. Informat. Rech. Opér. Anal. Numer.*, 12(4):325–334, 1978.

19. F. Girosi. An Equivalence Between Sparse Approximation and Support Vector Machines. *Neural Computation*, 10 (8):1455–1480, 1998.

20. K. Jetter, J. Stöckler, and J.D. Ward. Norming sets and scattered data approximation on spheres. In *Approximation Theory IX, Vol. II: Computational Aspects*, pages 137 – 144. Vanderbilt University Press, 1998.

21. K. Jetter, J. Stöckler, and J.D. Ward. Error estimates for scattered data interpolation on spheres. *Mathematics of Computation*, 68:733–747, 1999.

22. W. R. Madych. An estimate for multivariate interpolation II. *Journal of Approximation Theory*, 142:116–128, 2006.

23. C.A. Micchelli and M. Pontil. Learning the kernel function via regularization. *Journal of Machine Learning Research*, 6:1099–1125, 2005.

24. F. J. Narcowich, J. D. Ward, and H. Wendland. Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. *Mathematics of Computation*, 74:743–763, 2005.

25. C. Rieger. *Sampling Inequalities and Applications*. PhD thesis, Universität Göttingen, 2008.

26. C. Rieger and B. Zwicknagl. Deterministic error analysis of kernel-based regression and related kernel based algorithms. to appear in

Journal of Machine Learning Research, 2008.

27. C. Rieger and B. Zwicknagl. Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. to appear in Advances in Computational Mathematics, 2008.

28. R. Schaback. Recovery of functions from weak data using unsymmetric meshless kernel-based methods. *Applied Numerical Mathematics*, 58:726–741, 2007.

29. R. Schaback and H. Wendland. Inverse and saturation theorems for radial basis function interpolation. *Math. Comp.*, 71(238):669–681 (electronic), 2002.

30. R. Schaback and H. Wendland. Kernel techniques: from machine learning to meshless methods. *Acta Numerica*, 15:543–639, 2006.

31. B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, Cambridge, 2002.

32. B. Schölkopf, R. C. Williamson, and P. L. Bartlett. New Support Vector Algorithms. *Neural Computation*, 12:1207–1245, 2000.

33. V. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, New York, 1995.

34. G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1990.

35. H. Wendland. On the convergence of a general class of finite volume methods. *SIAM Journal of Numerical Analysis*, 43:987–1002, 2005.

36. H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2005.

37. H. Wendland and C. Rieger. Approximate interpolation with applications to selecting smoothing parameters. *Numer. Math.*, 101:643–662, 2005.

38. Z. Wu. Hermite–Birkhoff interpolation of scattered data by radial basis functions. *Approximation Theory and its Applications*, 8/2:1–10, 1992.

39. B. Zwicknagl. Power series kernels. *Constructive Approximation*, 29(1):61–84, 2009.